Clarification Potential of Instructions

Luciana Benotti
TALARIS Team - LORIA (Université Henri Poincaré, INRIA)
BP 239, 54506 Vandoeuvre-lès-Nancy, France
Luciana.Benotti@loria.fr

Abstract

Our hypothesis is that conversational implicatures are a rich source of clarification questions. In this paper we do two things. First, we motivate the hypothesis in theoretical, practical and empirical terms. Second, we present a framework for generating the clarification potential of an instruction by inferring its conversational implicatures with respect to a particular context. General means-ends inference, beyond classical planning, turns out to be crucial.

1 Introduction

Practical interest in clarification requests (CRs) no longer needs to be awakened in dialogue system designers (Gabsdil, 2003; Purver, 2004; Rodríguez and Schlangen, 2004; Rieser and Moore, 2005; Skantze, 2007). In sociolinguistics and discourse analysis, repair has been an even more favored theme for almost three decades now; see (Schegloff, 1987) as a representative example. However, the theoretical scope of the phenomena and its implications for a theory of meaning are still being delineated. Recently, it has been proposed that clarification should be a basic component in an adequate theory of meaning:

*The basic criterion for adequacy of a theory of meaning is the ability to characterize for any utterance type the **update** that emerges in the aftermath of successful mutual understanding and the full range of **possible clarification requests** otherwise — this is the early 21st century analogue of truth conditions. (Ginzburg, 2009, p.4)*

In this view, repairs are not a necessary evil but an intrinsic mechanism of language. In fact, interpreting an utterance centrally involves characterizing **the space of possible requests of clarification** of the utterance, that is its **clarification potential**. We believe that Ginzburg's comment points in the right direction; we discuss the motivations from a theoretical perspective in Section 2.1. In Section 2.2 we review a state-of-the-art definition of the notion of clarification from the perspective of dialogue system designers. This review makes evident the necessity of further refining the notion of clarification if it is going to play such a central role in a theory of meaning. In Section 2.3 we present our findings in the corpus SCARE (Stoia et al., 2008) which empirically motivates our work.

We believe that it is crucial to redefine the notion of clarification in functional terms. Because we know that the task is difficult, we restrict ourselves to one utterance type, **instructions**, and to a particular interaction level, the **task-level**. In the rest of the paper (Sections 3 and 4), we present a framework that generates the task-level clarification potential of an instruction by inferring its particularized conversational implicatures.

The following exchange illustrate the kinds of interactions our framework models:

(1) A(1): Turn it on.
B(2): By pushing the red button?
(Rodríguez and Schlangen, 2004, p.102)

Roughly speaking, our framework takes as input sentences like A(1) and explains how B(2) can be generated. In particular, the framework indicates what kinds of information resources and what kind of inferences are involved in the process of generating utterances like B(2). In other words, the goal of the framework is to explain why A(1) and B(2) constitute a coherent dialogue by saying how B(2) is relevant to A(1).

2 Background and motivation

In this section, we motivate our framework from the **theoretical perspective** of pragmaticists interested in the relevance of clarifications for a theory of meaning, from the **practical perspective** of dialogue system designers, and from the **empirical perspective** of a human-human corpus that provides evidence for the necessity of such a framework.

2.1 Theoretical: Relevance of clarifications

Modeling how listeners draw inferences from what they hear, is a basic problem for theories of understanding natural language. An important part of the information conveyed is inferred in context, given the nature of conversation as a goal-oriented enterprise; as illustrated by the following classical example by Grice:

(2) A: I am out of petrol.
B: There is a garage around the corner.
$\rightsquigarrow$ B thinks that the garage is open.
(Grice, 1975, p.311)

B's answer *conversationally implicates* ($\rightsquigarrow$) information that is relevant to A. In Grice's terms, B made a relevance implicature: he would be flouting the conversational maxim of relevance unless he believes that it's possible that the garage is open. A conversational implicature (CI) is different from an entailment in that it is *cancelable* without contradiction. B can append material that is inconsistent with the CI — "but I don't know whether it's open". Since the CI can be canceled, B knows that it does not necessarily hold and then both B or A are able to *reinforce* or *clarify* it without repetition.

It is often controversial whether something is actually a CI or not (people have different intuitions, which is not surprising given that people have different background assumptions). In dialogue, CRs provide good evidence of the implicatures that have been made simply because they make implicatures explicit. Take for instance the clarification request which can naturally follow Grice's example.

(3) A: and you think it's open?

B will have to answer and support the implicature (for instance with "yes, it's open till midnight") if he wants to get it added to the common ground; otherwise, if he didn't mean it, he can well reject it without contradiction with "well, you have a point there, they might have closed".

Our hypothesis is that CIs are a rich source of clarification requests. And our method for generating the potential CRs of an utterance will be then to infer (some of) the CIs of that utterance with respect to a particular context.

2.2 Practical: Kinds of clarifications

Giving a precise definition of a clarification request is more difficult than might be thought at first sight. Rodríguez and Schlangen (2004) recognize this problem by saying:

Where we cannot report reliability yet is for the task of identifying CRs in the first place. This is not a trivial problem, which we will address in future work. As far as we can see, Purver, Ginzburg and Healey have not tested for reliability for doing this task either. (Rodríguez and Schlangen, 2004, p.107)

One of the most developed classifications of CRs is the one presented in (Purver, 2004). However, Purver's classification relies mainly on the surface form of the CRs. The attempts found in the literature to give a classification of CRs according to their functions (Rodríguez and Schlangen, 2004; Rieser and Moore, 2005) are based on the four-level model of communication independently developed by Clark (1996) and Allwood (1995). The model is summarized in Figure 1 (from the point of view of the hearer).

Level	Clark	Allwood
4	consideration	reaction
3	understanding	understanding
2	identification	perception
1	attention	contact

Figure 1: The four levels of communication

Most of the previous work on clarifications has concentrated on levels 1 to 3 of communication. For instance, Schlangen (2004) proposed a fined-grained classification of CRs but only for level 3. Gabsdil (2003) proposes a test for identifying CRs. The test says that CRs cannot be preceded by explicit acknowledgements. But in the following example, presented by Gabsdil himself, the CR uttered by F can well start with an explicit "ok".

(4) G: I want you to go up the left hand side of it towards the green bay and make it a slightly diagonal line, towards, sloping to the right.
F: So you want me to go above the carpenter? (Gabsdil, 2003, p.30)

The kind of CR showed in 4, also called **clarification of intentions** or **task level clarifications**, are in fact very frequent in dialogue; they have been reported to be the second or third most common kind of CR (the most common being reference resolution). (Rodríguez and Schlangen, 2004) reports that 22% of the CRs found by them in a German task-oriented spoken dialogue belonged to level 4, while (Rieser and Moore, 2005) reports 8% (a high percentage considering that the channel quality was poor and caused a 31% of acoustic problems).

Fourth level CRs are not only frequent but there are studies that show that the hearer in fact prefers them. That is, if the dialogue shows a higher amount of task related clarifications (instead of, conventional CRs such as "what?") hearers qualitative evaluate the task as more successful (Skantze, 2007). (Gabsdil, 2003) and (Rieser and Moore, 2005) also agree that for task-oriented dialogues the hearer should present a task-level reformulation to be confirmed rather than asking for repetition, thereby showing his subjective understanding to the other dialogue participants. Gabsdil briefly suggests a step in this direction:

Task-level reformulations might benefit from systems that have access to effects of action operators or other ways to compute task-level implications. (Gabsdil, 2003, p.29 and p.34)

In the rest of the paper we propose a framework that formalizes how to compute task-level implicatures and that suggests a finer-grained classification for CRs in level 4. But first, in Section 2.3 we present empirical findings that motivate such a framework.

2.3 Empirical: The SCARE corpus

The SCARE corpus (Stoia et al., 2008) consists of fifteen English spontaneous dialogues situated in an instruction giving task[1]. It was collected using the Quake environment, a first-person virtual reality game. The task consists of a direction giver (DG) instructing a direction follower (DF) on how to complete several tasks in a simulated game world. The corpus contains the collected audio and video, as well as word-aligned transcriptions.

The DF had no prior knowledge of the world map or tasks and relied on his partner, the DG, to guide him on completing the tasks. The DG had a map of the world and a list of tasks to complete (detailed in Appendix A.3). The partners spoke to each other through headset microphones; they could not see each other. As the participants collaborated on the tasks, the DG had instant feedback of the DF's location in the simulated world, because the game engine displayed the DF's first person view of the world on both the DG's and DF's computer monitors.

We analyzed the 15 transcripts that constitute the SCARE corpus while watching the associated videos to get familiarized with the experiment and evaluate its suitability for our purposes. Then, we randomly selected one dialogue; its transcript contains 449 turns and its video lasts 9 minutes and 12 seconds. Finally, we classified the clarification requests according to the levels of communication (see Figure 1). We found 29 clarification requests; so 6.5% of the turns are CRs. From these 29 CRs, 65% belong to the level 4 of Table 1, and 31% belonged to level 3 (most of them related to reference resolution). Only 4% of the CRs were acoustic (level 2) since the channel used was very reliable.

In fact we only found one CR of the form "what?" and it was a signal of incredulity of the effect of an action as can be seen below:

DG(1): and then cabinet should open
DF(2): did it
DF(3): nothing in it
DG(4): what?
DG(5): There should be a silencer there

Interestingly, the "what?" form of CR was reported as the most frequently found in "ordinary" dialogue in (Purver et al., 2003). This is not the case in the SCARE corpus. Furthermore, "what?" is usually assumed to be a CR that indicates a low level of coordination and is frequently classified as belonging to level 1 or 2. However, this is not the case in our example in which the CR is evidently related to the task structure and thus belongs to level 4. This is an example of why surface form is not reliable when classifying CRs.

[1]The corpus is freely available for research in http://slate.cse.ohio-state.edu/quake-corpora/scare/

2.4 Preliminary conclusions

In this preliminary study, the SCARE corpus seems to present more CRs than the corpus analyzed by previous work (which reports that 4% of the dialogue turns are CR). Furthermore, in distinction to results reported in Ginzburg (2009), most CRs occur at level 4. We believe this is naturally explained in politeness theory (Brown and Levinson, 1987).

The participants were punished if they performed steps of the task that they were not supposed to (see the instructions in Appendix A.1). This punishment might take precedence over the dispreference for CRs that is universal in dialogue due to politeness. CRs are perceived as a form of disagreement which is universally dispreferred according to politeness theory. The pairs of participants selected were friends so the level of intimacy among them was high, lowering the need of politeness strategies; a behavior that is also predicted by politeness theory. Finally, the participants received a set of instructions before the task started (see Appendix A) that includes information on the available actions in the simulated world and their expected effects. The participants make heavy use of this to produce high level clarification requests, instead of just signaling misunderstanding.

From these observations we draw the preliminary conclusion that clarification strategies depend on the information that is available to the dialogue participants (crucially including the information available before the dialogue starts) and on the constraints imposed on the interaction, such as politeness constraints. In Section 3 we describe the four information resources of our framework whose content depends on the information available to the dialogue participants. In Section 4 we introduce the reasoning tasks that use the information resources to infer the clarification potential of instructions. The study of the interaction between politeness constraints and clarification strategies seems promising, and we plan to address it in future work.

3 The information resources

The inference framework uses four information resources whose content depends on the information available to the dialogue participants. We describe each of them in turn and we illustrate their content using the SCARE experimental setup.

3.1 The world model

Since the kind of utterance that the framework handles are instructions that are supposed to be executed in a simulated world, the first required information resource is a model of this world. The world model is a knowledge base that represents the physical state of the simulated world. This knowledge base has complete and accurate information about the world that is relevant for completing the task at hand. It specifies properties of particular individuals (for example, an individual can be a *button* or a *cabinet*). Relationships between individuals are also represented here (such as the relationship between an object and its location). Such a knowledge base can be thought as a first-order model.

The content of the world model for the SCARE setup is a representation of the factual information provided to the DG before the experiment started, namely, a relational model of the map he received (see Figure 3 in Appendix A.3). Crucially, such a model contains all the functions associated with the buttons in the world and the contents of the cabinets (which are indicated on the map).

3.2 The dialogue model

Usually, this knowledge base starts empty; it is assumed to represent what the DF knows about the world. The information learned, either through the contributions made during the dialogue or by navigating the simulated world, are incrementally added to this knowledge base. The knowledge is also represented as a relational model and in fact this knowledge base will usually (but not necessarily) be a submodel of the world model.

The DF initial instructions in the SCARE setup include almost no factual information (as you can verify looking at his instructions in Appendix A.2). The only factual information that he received were pictures of some objects in the world so that he is able to recognize them. Such information is relevant mainly for referent resolution and this is not the focus of the current paper. Therefore, for our purposes we can assume that the dialogue model of the SCARE experiment starts empty.

3.3 The world actions

Crucially, the framework also includes the definitions of the actions that can be executed in the world (such as the actions *take* or *open*). Each ac-

tion is specified as a STRIPS-like operator (Fikes et al., 1972) detailing its arguments, preconditions and effects. The preconditions indicate the conditions that the world scenario must satisfy so that the action can be executed; the effects determine how the action changes the world when it is executed. These actions specify complete and accurate information about how the world behaves and together with the world model is assumed to represent what the DG knows about the world.

The SCARE world action database will contain a representation of the specification of the quake controls (see Appendix A.1) received by both participants and the extra action information that the DG received. First, he received a specification of the action *hide* that was not received by the DF. Second, if the DG read the instructions carefully, he knows that pressing a button can also cause things to move. The representation of this last action schema is shown in Appendix A.3.1.

3.4 The potential actions

The potential actions include representation of actions that the DF learned from the instructions he received before beginning the task. This includes the quake controls (see Appendix A.1) and also the action knowledge that he acquired during his learning phase (see appendix A.2). In the learning phase the direction follower learned that the effect of pressing a button can open a cabinet (if it was closed) or close it (if it was opened). Such knowledge is represented as a STRIPS-like operator like one showed in Appendix A.2.1.

3.5 Preliminary conclusions

An **action language** like PDDL (Gerevini and Long, 2005) can be used to specify the two action databases introduced above (in fact, the STRIPS fragment is enough). PDDL is the official language of the International Conference on Automated Planning and Scheduling since 1998. This means that most off-the-shelf planners that are available nowadays support this language, such as FF (Hoffmann and Nebel, 2001) and SGPlan (Hsu et al., 2006).

As we said in the previous section, the world model and the dialogue model are just relational structures like the one showed in Figure 3 (in the appendix). These relational structures can be directly expressed as a set of literals which is the format used to specify the **initial state** of a planning problem.

The information resources then constitute almost everything that is needed in order to specify a complete **planning problem**, as expected by current planners, the only element that the framework is missing is the **goal**. With a set of action schemas (i.e. action operators), an initial state and a goal as input, a planner is able to return a sequence of actions (i.e. a plan) that, when executed in the initial state, achieves the goal.

Planning is a **means-end inference task**, a kind of **practical inference** as defined by Kenny in (Kenny, 1966); and is a very popular inference task indeed as evidenced by the amount of work done in the area in the last two decades. However, *planning is not the only interesting means-end inference task*. One of the goals of the next section is to show exactly this: there is more to practical inference than planning.

4 The inference tasks

In this section we do two things. First, we say how current off-the-shelf planners can be used to infer part of the clarification potential of instructions. In particular we define what the missing element, the goal, is and we illustrate this with fragments of human-human dialogue of the SCARE corpus. Incidentally, we also show that clarification potential can not only be used for generating and interpreting CRs but also for performing acceptance and rejection acts. Second, we motivate and start to define one means-ends inference task that is not currently implemented, but that is crucial for inferring the clarification potential of instructions.

In order to better understand the examples below you may want to read the Appendix A first. The information in the Appendix was available to the participants when they performed the experiments and it's heavily used in the inferences they draw.

4.1 Planning: A means-end inference task

Shared-plan recognition —and *not* artificial intelligence planning— has been used for utterance interpretation (Lochbaum, 1998; Carberry and Lambert, 1999; Blaylock and Allen, 2005). In such plan recognition approaches each utterance adds a constraint to the plan that is partially filled out, and the goal of the conversation has to be inferred during the dialogue; that is, a *whole dialogue* is mapped to one shared plan. In our approach, *each instruction* is interpreted as a plan instead; that is,

we use planning at the utterance level and not at dialogue level.

Artificial intelligence planning has been used at utterance level (called micro-planning) for *generation* (Koller and Stone, 2007). We use artificial intelligence planning for *interpretation* of instructions instead.

In our framework, the goal of the planning problem are the *preconditions of instruction* for which the clarification potential is being calculated. Now, the planning problem has a goal, but there are two action databases and two initial states. Which one will be used for finding the clarification potential? In fact, all four.

When the DG gives an instruction, the DF has to interpret it in order to know what actions he has to perform (step 1 of the inference). The interpretation consists in trying to construct a plan that, when executed in the current state of the game world, achieves the goals of the instruction. The specification of such planning problem is as follows. The preconditions of the instruction are the *goal* of the planning problem, the dialogue model is the *initial state* and the potential actions are the *action operators*. With this information the off-the-shelf planner will find a *plan*, a sequence of actions that are the implicatures of the instruction.

Then (step 2 of the inference), an attempt to execute the plan on the the world model and using the world actions occurs. *Whenever the plan fails, there is a potential clarification.*

Using clarification potential to clarify: In the dialogue below, the participants are trying to move a picture from a wall to another wall (task 1 in Appendix A.3). The instruction that is being interpreted is the one uttered by the DG in (1). Using the information in the potential action database, the DF infers a plan that involves two implicatures, namely *picking up the picture* (in order to achieve the precondition of holding the picture), and *going to the wall* (inference step 1). However, this plan will fail when executed on the world model because the picture is *not takeable* and thus it cannot be picked, resulting in a potential clarification (inference step 2). This potential clarification, foreshadowed by (3), is finally made explicit by the CR in (4).

DG(1): well, put it on the opposite wall
DF(2): ok, control picks the .
DF(3): control's supposed to pick things up and .
DF(4): am I supposed to pick this thing?

A graphical representation of both steps of inference involved in this example is shown in Section B of the Appendix[2].

But also to produce evidence of rejection: In the dialogue below, the DG utters the instruction (1) knowing that the DF will not be able to follow it; the DG is just thinking aloud. If taken seriously, this instruction would involve the action *resolve the reference "cabinet nine"*. A precondition of this action is that the DF knows the numbers of the cabinets, but both participants know this is not the case, only the DG can see the map. That's why the rejection in (2) is received with laughs and the DG continues his loud thinking in (3) while looking at the map.

DG(1): we have to put it in cabinet nine .
DF(2): yeah, they're not numbered [laughs]
DG(3): [laughs] where is cabinet nine .

And to produce evidence of acceptance: The following dialogue fragment continues the fragment above. Now, the DG finally says where cabinet nine is in (4). And the DF comes up with the plan that he incrementally grounds making it explicit in (5), (7), and (9) while he is executing it; the plan achieves the precondition of the instruction *put* of being near the destination of the action, in this case "near cabinet nine". Uttering the steps of the plan that were not made explicit by the instruction is indeed a frequently used method for performing acceptance acts.

DG(4): it's . kinda like back where you started . so
DF(5): ok . so I have to go back through here .
DG(6): yeah
DF(7): and around the corner .
DG(8): right
DF(9): and then do I have to go back up the steps
DG(10): yeah
DF(11): alright, this is where we started
DG(12): ok . so your left ca- . the left one
DF(13): alright, so how do I open it?

In (13) the DF is not able to find a plan that achieves another precondition of the action *put*, namely that the destination container is opened, so he directly produces a CR about the precondition.

[2]The correct plan to achieve (1) involves pressing button 12, as you (and the DG) can verify on the map (in the Appendix).

4.2 Beyond classical planning: Other important means-end inference tasks

Consider the following example, here the DG just told the DF to press a button, in turn (1), with no further explanation. As a result of the action a cabinet opened, and the DF predicted that the following action requested would be (5). In (6) the DG confirms this hypothesis.

DG(1): press the button on the left [pause]
DG(2): and . uh [pause]
DF(3): [pause]
DG(4): [pause]
DF(5): put it in this cabinet?
DG(6): put it in that cabinet, yeah

The inference that the DF did in order to produce (5) can be defined as another means-end inference task which involves finding the **next relevant actions**. The input of such task would also consist of an initial state, a set of possible actions but it will contain one observed action (in the example, action (1)). Inferring the next relevant action consists in inferring the affordabilities (i.e. the set of executable actions) of the initial state and the affordabilities of the state after the observed action was executed. The **next relevant actions** will be those actions that were activated by the observed action. In the example above, the next relevant action that will be inferred is "put the thing you are carrying in the cabinet that just opened", just what the DF predicted in (5).

The definition of this inference task needs refining but it already constitutes an interesting example of a new form of means-ends reasoning.

There are further examples in the corpus that suggest the need for means-end inferences in situations in which a classical planner would just say "there is no plan". These are cases in which no complete plan can be found but the DF is anyway able to predict a possible course of action. For instance, in the last dialogue of Section 4.1, the DF does not stops in (13) and waits for an answer but he continues with:

DF(14): one of the buttons?
DG(15): yeah, it's the left one

Other CRs similar to this one, where a parameter of the action is ambiguous, is missing or is redundant, were also found in the corpus.

4.3 Preliminary Conclusions

The inference-tasks we discussed or just hinted to in this paper do not give a complete characterization of the kinds of clarification requests of level 4. It covers 14 of the 19 CRs in the SCARE dialogue analyzed in Section 2.3. CRs not covered at all have to do mainly with the fact that people do not completely remember (or trust) the instructions during the experiments or what themselves (or their partner) said a few turns before, such as the following one:

DG(1): you've to . like jump on it or something .
DF(2): I don't know if I can jump

Here, the DF does not remember that he can jump using the Spacebar as stated in the instructions he received (Appendix A.1).

In order to account for these cases it is necessary to consider how conversation is useful for overcoming also this issue. The fact that people's memory is non reliable is intrinsic to communication and here again, communication must provide intrinsic mechanisms to deal with it. Modeling such things are challenges that a complete theory of communication will have to face.

5 Conclusions

Conversational implicatures are negotiable, this is the characteristic that distinguishes them from other kinds of meanings (like entailments). Dialogue provides an intrinsic mechanism for carrying out negotiations of meaning, namely clarifications. So our hypothesis is that conversational implicatures are a rich source of clarification requests.

In order to investigate this hypothesis, we reviewed theoretical work from pragmatics, practical work from the dialogue system community and we presented empirical evidence from spontaneous dialogues situated in an instruction giving task. Also, we presented a framework in which (part of) the clarification potential of an instruction is generated by inferring its conversational implicatures. We believe that this is a step towards defining a clear functional criteria for identifying and classifying the clarification requests at level 4 of communication.

But much more remains to be done. The empirical results we present here are suggestive but preliminary; we are currently in the process of evaluating their reliability measuring inter-annotator

agreement. Moreover, in the course of this work we noticed a promising link between clarification strategies and politeness constraints which we plan to develop in future work. Also, we are particularly interested in means-ends reasoning other than planning, something we have merely hinted at in this paper; these tasks still need to be formally defined, implemented and tested. Finally, we are considering the GIVE challenge (Byron et al., 2009) as a possible setting for evaluating our work (our framework could predict potential clarification requests from the users).

There is lot to do yet, but we believe that the interplay between conversational implicatures and clarification mechanisms will play a crucial role in future theories of communication.

References

Jens Allwood. 1995. An activity based approach to pragmatics. In *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics*, pages 47–80. University of Göteborg.

Nate Blaylock and James Allen. 2005. A collaborative problem-solving model of dialogue. In *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*, pages 200–211, Lisbon, Portugal.

Penelope Brown and Stephen Levinson. 1987. *Politeness: Some universals in language usage*. Studies in Interactional Sociolinguistics.

Donna Byron, Alexander Koller, Kristina Striegnitz, Justine Cassell, Robert Dale, Johanna Moore, and Jon Oberlander. 2009. Report on the First NLG Challenge on Generating Instructions in Virtual Environments (GIVE). In *Proc. of the 12th European Workshop on Natural Language Generation*, pages 165–173, Athens, Greece. ACL.

Sandra Carberry and Lynn Lambert. 1999. A process model for recognizing communicative acts and modeling negotiation subdialogues. *Computational Linguistics*, 25(1):1–53.

Herbert Clark. 1996. *Using Language*. Cambridge University Press, New York.

Richard Fikes, Peter Hart, and Nils Nilsson. 1972. Learning and executing generalized robot plans. *Artificial Intelligence*, 3:251–288.

Malte Gabsdil. 2003. Clarification in spoken dialogue systems. In *Proc of the AAAI Spring Symposium. Workshop on Natural Language Generation in Spoken and Written Dialogue*, pages 28–35.

Alfonso Gerevini and Derek Long. 2005. Plan constraints and preferences in PDDL3. Technical Report R.T. 2005-08-47, Brescia University, Italy.

Jonathan Ginzburg. 2009. *The interactive Stance: Meaning for Conversation*. CSLI Publications.

Paul Grice. 1975. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics: Vol. 3: Speech Acts*, pages 41–58. Academic Press.

Jörg Hoffmann and Bernhard Nebel. 2001. The FF planning system: Fast plan generation through heuristic search. *JAIR*, 14:253–302.

Chih-Wei Hsu, Benjamin W. Wah, Ruoyun Huang, and Yixin Chen. 2006. New features in SGPlan for handling soft constraints and goal preferences in PDDL3.0. In *Proc of ICAPS*.

Anthony Kenny. 1966. Practical inference. *Analysis*, 26:65–75.

Alexander Koller and Matthew Stone. 2007. Sentence generation as planning. In *Proc. of ACL-07*, Prague.

Karen E. Lochbaum. 1998. A collaborative planning model of intentional structure. *Comput. Linguist.*, 24(4):525–572.

Matthew Purver, Jonathan Ginzburg, and Patrick Healey. 2003. On the means for clarification in dialogue. In *Current and New Directions in Discourse and Dialogue*, pages 235–255. Kluwer Academic Publishers.

Matthew Purver. 2004. *The Theory and Use of Clarification Requests in Dialogue*. Ph.D. thesis, King's College, University of London.

Verena Rieser and Johanna Moore. 2005. Implications for generating clarification requests in task-oriented dialogues. In *Proc of ACL*, pages 239–246.

Kepa Rodríguez and David Schlangen. 2004. Form, intonation and function of clarification requests in german task oriented spoken dialogues. In *Proc of SEMDIAL*, pages 101–108.

Emanuel Schegloff. 1987. Some sources of misunderstanding in talk-in-interaction. *Linguistics*, 8:201–218.

David Schlangen. 2004. Causes and strategies for requesting clarification in dialogue. In *Proc of SIGDIAL*.

Gabriel Skantze. 2007. *Error Handling in Spoken Dialogue Systems*. Ph.D. thesis, KTH - Royal Institute of Technology, Sweden.

Laura Stoia, Darla Shockley, Donna Byron, and Eric Fosler-Lussier. 2008. SCARE: A situated corpus with annotated referring expressions. In *Proc of LREC*.

Laura Stoia. 2007. *Noun Phrase Generation for Situated Dialogs*. Ph.D. thesis, Ohio State University, USA.

A Instructions for the DG and DF

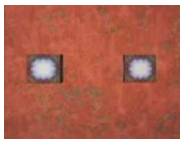
In this section, we specify the information that was available to the DG and the DF before the SCARE experiment started (adapted from (Stoia, 2007)). These instructions are crucial for our study since they define the content of the information resources of the inference framework described in this paper.

A.1 Instructions for both

The following specification of the Quake controls, that is, the possible actions in the simulated world, were received by all participants.

1. Use the arrow keys for **movement**:
   - Walk forward: ↑
   - Walk backward: ↓
   - Turn right: →
   - Turn left: ←
2. To **jump**: use Spacebar.
3. To **press a button**: Walk over the button. You will see it depress.
4. To **pick up an object**: Step onto the item then press Ctrl (Control key).
5. To **drop an object**: Hit TAB to see the list of items that you are currently carrying. Press the letter beside the item you wish to drop. Press TAB again to make the menu go away.

The participants also received the following pictures of possible objects in the simulated world so that they are able to recognize them.

Buttons	Cabinet
	

The following things were indicated as being objects that the DF can pick up and move:

Quad damage	Rebreather	Silencer
		

They also received the following warning: You will not be timed, but penalty points will be taken for pushing the wrong buttons or placing things in the wrong cabinets.

A.2 Instructions for the Direction Follower

Only the DF received the following information:

Phase 1: Learning the controls First you will be put into a small map with no partner, to get accustomed to the quake controls (detailed in Section A.1). Practice moving around using the arrow keys. Practice these actions:

1. Pick up the Rebreather or the Quad Damage.
2. Push the blue button to open the cabinet.
3. Drop the Quad Damage or the Rebreather inside the cabinet and close the door by pushing the button again.

Phase 2: Completing the task In this phase you will be put in a new location. Your partner will direct you in completing 5 tasks. He will see the same view that you are seeing, but you are the only one that can move around and act in the world.

A.2.1 Implications for the Potential Actions

In phase 1, when the DF is learning the controls, he learns that buttons can have the effect of opening closed cabinets and closing open cabinets. Such action is formalized as follows in PDDL (Gerevini and Long, 2005) and is included in the possible action database:

```
(:action press_button
   :parameters (?x ?y)
   :precondition
        (button ?x)
        (cabinet ?y)
        (opens ?x ?y)
   :effects
        (when (open ?y) (closed ?y))
        (when (closed ?y) (open ?y)))
```

Notice that this action operator has conditional effects in order to specify the action more succinctly. However, it is not mandatory for the action language to support conditional effects. This action could be specified with two actions in which the antecedent of the conditional effect is now a precondition.

A.3 Instructions for the Direction Giver

Only the DG received the following information:

Phase 1: Planning the task Your packet contains a **map** of the quake world with **5 objectives** that you have to direct your partner to perform. Read the instructions and take your time to plan the directions you want to give to your partner.

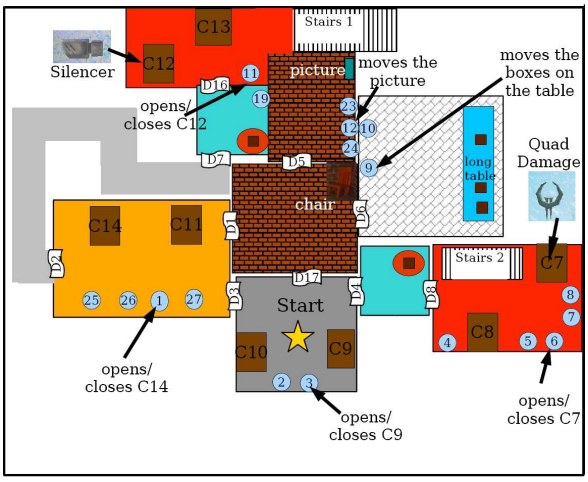


Figure 2: Map received by the DG (upper floor)

Phase 2: Directing the follower In this phase your partner will be placed into the world in the start position. Your monitor will show his/her view of the world as he/she moves around. He/she has no knowledge of the tasks, and has not received a map. You have to direct him/her through speech in order to complete the tasks. The objective is to complete all 5 tasks, but the order does not matter.

The tasks are:

1. Move the picture to the other wall.
2. Move the boxes on the long table so that the final configuration matches the picture below.

Picture	Long table
	

3. Hide the Rebreather in Cabinet9. To **hide** an item you have to find it, pick it up, drop it in the cabinet and close the door.
4. Hide the Silencer in Cabinet4.
5. Hide the Quad Damage in Cabinet14.
6. At the end, return to the starting point.

A.3.1 Implications for the World Actions

The functions of the buttons that can move things can be represented in the following action schema. If the thing is in it's original location (its location when the game starts), we say that is thing is `not-moved`. If the thing is in the goal position then we say that the thing is `moved`.

```
(:action press_button
   :parameters (?x ?y)
   :precondition
        (button ?x)
        (thing ?y)
        (moves ?x ?y)
   :effects
        (when (moved ?y) (not-moved ?y))
        (when (not-moved ?y) (moved ?y)))
```

A.3.2 Implications for the World Model

The world model is a relational model that represents the information provided by the map, including the functions of the buttons and the contents of the cabinets.

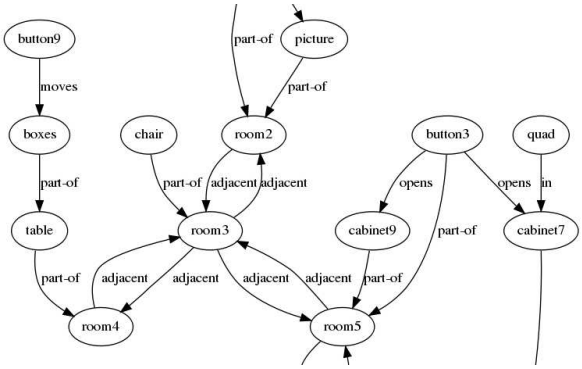


Figure 3: Fragment of the SCARE world model

B Clarification Potential Inference Steps

The following pictures illustrate how the implicatures of the instruction "put the picture on the opposite wall" are calculated using the dialogue model (Figure 4) and used to predict the CR "Am I supposed to pick up this thing?" (Figure 5).

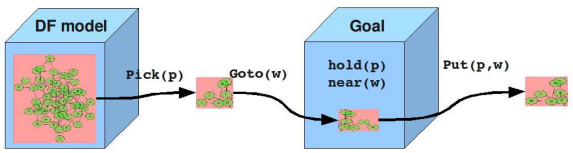


Figure 4: Step 1 - Calculating the implicatures

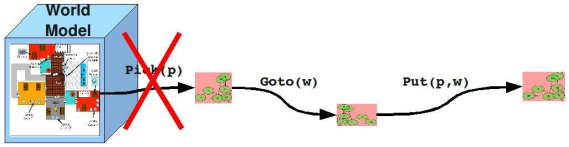


Figure 5: Step 2 - Predicting the CR